PB-218 503

ESTIMATIONS OF THE PARAMETERS OF THE LAW OF EXTREME VALUES AND ECONOMIC CONSEQUENCES

Neil Goldstein

Columbia University
New York, New York

1966

DISTRIBUTED BY:

NTIS

National Technical Information Service
U. S. DEPARTMENT OF COMMERCE
5285 Port Royal Road, Springfield Va. 22151

WP-00457-01

PB 218 503

ESTIMATIONS OF THE PARAMETERS OF THE LAW OF EXTREME VALUES AND ECONOMIC CONSEQUENCES.

ESTIMATIONS DES PARAMÈTRES DE LA LOI DES VALEURS EXTRÊMES ET CONSÉQUENCES ÉCONOMIQUES

Neil GOLDSTEIN[1]
(U.S. Public Health Service)

L'ingénieur étudie et construit un ouvrage (barrage, pont,...) en vue de sa durée. Pour cela, il doit connaître, ou tout au moins être capable d'estimer l'importance des risques aléatoires auxquels, dans l'avenir, est exposé son ouvrage. Cela signifie qu'il doit savoir quelle peut être l'importance d'une crue, la force d'un vent... auxquelles son ouvrage risque d'être soumis. Pour cela, l'ingénieur a à sa disposition la théorie statistique des valeurs extrêmes.

Afin de mettre cette théorie statistique en pratique, certaines estimations doivent être faites : le problème étudié par l'ingénieur est, dans une certaine mesure, réduit à un problème d'estimation, il s'agit alors de choisir parmi les estimations possibles des paramètres d'une distribution celles qu'il considérera comme les meilleures.

Dans le cas où l'ingénieur doit tenir compte de phénomènes dont la résultante, dans le cas de destruction de son ouvrage, peut être la perte de vies humaines ou un coût de remplacement très élevé (comparé au coût initial) il sera certainement plus particulièrement intéressé par le nombre, pouvant être envisagé, des ouvrages détruits au cours d'une période donnée (c'est-à-dire la période d'usage prévue dans son projet). Ceci nous conduit à définir une procédure d'estimation comme étant meilleure qu'une autre, si elle conduit à une estimation plus petite du nombre de dépassements des valeurs extrêmes envisagées ou des ouvrages détruits.

La distribution généralement utilisée pour l'analyse du débit des crues, de la vitesse du vent, ou d'autres phénomènes analogues, est la première distribution asymptotique des valeurs extrêmes dont la fonction de répartition est :

$$\Phi(x) = \exp. - [e^{-\alpha(x-u)}] \qquad (1)$$

où x est une variable aléatoire pouvant varier de $-\infty$ à $+\infty$ et α et u sont les paramètres de la distribution qui doivent être estimés.

---

(1) Cet article écrit comme assistant au professeur E.J. Gumbel à Columbia University et sous les auspices de U.S. Public Health Service (Grant WP 00457-01), fait suite aux publications qui ont parues dans la Revue de Statistique Appliquée, vol. 5, n° 2, et n° 4, 1957 ; vol. 10, n° 1, 1962. Le travail présent essaye d'appliquer la théorie statistique des valeurs extrêmes à la solution de quelques problèmes économiques qui se présentent dans les constructions hydroélectriques.

Reproduced by
NATIONAL TECHNICAL
INFORMATION SERVICE
U S Department of Commerce
Springfield VA 22151

Il existe actuellement deux méthodes qui sont les plus souvent utilisées pour l'estimation de α et u.

Le but de cette étude est de comparer ces deux méthodes afin de voir laquelle est la meilleure du point de vue défini ci-dessus.

L'une de ces deux méthodes est la méthode classique des moments, la seconde en étant une modification proposée par Gumbel. La comparaison de ces deux méthodes a été souvent étudiée de divers points de vue (cf. par exemple Lieblein [1]). Une troisième méthode est celle du maximum de vraisemblance : bien que cette méthode présente des avantages théoriques, elle augmente notablement les difficultés de calcul. La prévision obtenue par cette méthode est d'ailleurs, en générale très voisine de celle donnée par la méthode des moments.

La méthode classique des moments conduit aux estimations :

$$\frac{1}{\alpha_1} = \frac{\sqrt{6}}{\pi} s \tag{2}$$

et :

$$\hat{u}_1 = \bar{x} - \gamma \frac{\sqrt{6}}{\pi} s \tag{3}$$

où $\bar{x}$ est la moyenne de l'échantillon, s l'écart-type de l'échantillon et $\gamma = 0,57722 \ldots$ est la constante d'Euler.

Les estimations modifiées [2] sont données par :

$$\frac{1}{\alpha_2} = \frac{s}{\sigma_n} \tag{4}$$

et :

$$\hat{u}_2 = \bar{x} - \frac{\bar{y}_n}{\sigma_n} s \tag{5}$$

avec :

$$\bar{y}_n = \frac{1}{n} \sum_{m=1}^{n} \left\{ - \ln \left[- \ln \left(\frac{m}{n+1} \right) \right] \right\}$$

et :

$$\sigma_n = [\bar{y}_n^2 - (\bar{y}_n)^2]^{1/2}$$

Des tables ont été calculées pour ces deux valeurs [3]. Lorsque l'effectif n de l'échantillon augmente, les estimations modifiées se rapprochent de celles fournies par la méthode des moments : dans un but de comparaison, la signification des différences va être étudiée seulement pour de petits échantillons.

La formule (1) peut aussi être écrite sous forme non paramétrique si l'on fait la transformation linéaire :

$$y = \alpha(x - u) \tag{6}$$

La variable y est appelée variable réduite. Elle a la propriété, comme toute variable réduite de condenser les observations. En fait, dans la plupart des cas pratiques, y reste dans l'intervalle $-2 < y < 7$.

Renversant la procédure, si l'on part avec des valeurs extrêmes réduites y et si l'on fait la transformation inverse :

$$x = u + y/\alpha \tag{7}$$

on obtient des valeurs "non réduites", équivalentes aux observations réelles.

Cette transformation inverse aura pour effet d'augmenter l'étendue et par conséquent la variance des observations.

Il en résulte qu'une transformation de cette espèce augmentera, dans la plupart des cas, les plus grandes valeurs cependant qu'elle diminuera les plus petites.

De plus, le nombre total des dépassements sera augmenté, puisque la diminution des valeurs plus petites sera légère comparée au pourcentage correspondant d'augmentation des valeurs plus grandes.

Le problème est alors de trouver la transformation qui augmentera aussi peu que possible le nombre total des dépassements : ceci sera obtenu en augmentant plus largement les valeurs extrêmes.

Une prévision des crues considérée comme fonction de la durée de retour T (prédéterminée par l'ingénieur dans son projet) et obtenue à partir de la méthode des moments en utilisant les équations (2), (3) et (7) est :

$$\hat{x}_1(T) = \bar{x} + \frac{\sqrt{6}}{\pi} s (y_T - \gamma) \tag{8}$$

De même, la méthode modifiée, à partir des équations (4), (5) et (7) donnera :

$$\hat{x}_2(T) = \bar{x} + \frac{1}{\sigma_n} s (\bar{y}_T - \bar{y}_n) \tag{9}$$

Ces deux prévisions seront égales et les deux droites représentatives se couperont pour :

$$y_T = \frac{\frac{\sqrt{6}}{\pi} \gamma \sigma_n - \bar{y}_n}{\frac{\sqrt{6}}{\pi} \sigma_n - 1} \tag{10}$$

Cette valeur de y_T se trouve dans l'intervalle $0 < y_T < \bar{y}_n < \gamma$. Pour $100 > n > 10$ on aura $0,261 < y_T < 0,287$ ou, de manière équivalente sur l'échelle des probabilités, $0,463 < \Phi(y_T) < 0,472$. Il en résulte que pour les durées de retour plus grandes qu'environ 1,86, cas les plus intéressants dans un but d'extrapolation, on aura $\hat{x}_2 > \hat{x}_1$, $\hat{x}_2$ augmentant plus rapidement que $\hat{x}_1$: la méthode modifiée conduit donc à un nombre total plus petit de dépassements. Ceci signifie que la méthode modifiée incite l'ingénieur à construire son ouvrage pour qu'il résiste à des crues plus importantes que celles correspondant à l'autre méthode. D'autre part, on récolte ainsi, avec la méthode modifiée, la moisson d'économies financières lorsque les coûts de réparation et de remplacement sont pris en considération.

En vue de démontrer ces principes et de rechercher les ordres de grandeur qu'ils entraînent on a utilisé la méthode de Monte Carlo : recherche de la solution de problèmes mathématiques se posant dans un

contexte aléatoire à l'aide d'expériences par échantillonnage. La méthode consiste à construire un modèle stochastique artificiel du processus mathématique et à prélever des échantillons dans ce modèle.

Il est admis comme hypothèse initiale que l'on doit interpréter des valeurs extrêmes au hasard comme étant les plus grandes valeurs dont l'ingénieur peut craindre l'effet sur son ouvrage, c'est-à-dire que si un ensemble fini de valeurs est tiré au hasard de la distribution définie par (1), le nombre total de dépassements représente le nombre total de ponts emportés ou de barrages détruits au cours d'une période de temps spécifiée (une interprétation analogue a été donné par Benson [4] qui utilisa une population de 1000 valeurs distribuée conformément à (1) et appliqua ensuite une procédure d'échantillonnage pour comparer l'influence de diverses procédures.

Une série de 500 valeurs extrêmes [5], dont on avait préalablement vérifié qu'elle satisfaisait aux tests classiques du caractère aléatoire, a été utilisée comme base des valeurs extrêmes réduites, y. Les 500 valeurs extrêmes ont été d'abord partagées en 50 sous-échantillons consécutifs d'effectif 10 et les deux paramètres ont été calculés pour ces 50 sous-échantillons à l'aide des formules (2), (3), (4) et (5) [6] (Tableau 1).

Tableau 1

Estimation des paramètres par les deux méthodes

	Méthode des moments						Méthode des moments modifiés.				
	$1/\hat{\alpha}_1$	$\hat{u}_1$		$1/\hat{\alpha}_1$	$\hat{u}_1$		$1/\hat{\alpha}_1$	$\hat{u}_2$		$1/\hat{\alpha}_1$	$\hat{u}_2$
1	.8203	-.0098	26	.9233	.0603	1	1.1078	-.0850	26	1.2469	-.0242
2	.8096	-.2754	27	.9196	.1726	2	1.0934	-.3495	27	1.2420	.0884
3	1.0061	.2062	28	.7618	-.4829	3	1.3588	.1141	28	1.0288	-.5526
4	.8146	.0722	29	1.0404	.2863	4	1.1001	-.0023	29	1.4051	.1910
5	.8011	-.0957	30	1.1644	-.5043	5	1.0819	-.1690	30	1.5726	-.6109
6	.9489	-.1249	31	.6506	.0744	6	1.2815	-.2118	31	.8787	.0148
7	1.0162	.0300	32	1.2230	-.2590	7	1.3724	-.0630	32	1.6517	-.3710
8	1.1148	-.3192	33	1.2149	.2319	8	1.5056	-.4213	33	1.6408	.1206
9	1.2158	.0010	34	1.0930	.1463	9	1.6420	-.1103	34	1.4762	.0462
10	.8618	.0023	35	1.1955	-.0521	10	1.1639	-.0766	35	1.6146	-.1616
11	.8467	.5589	36	1.2192	-.1164	11	1.1435	.4814	36	1.6466	-.2281
12	1.3172	-.0479	37	.8264	.1892	12	1.7789	-.1886	37	1.1161	.1136
13	1.1035	-.1591	38	.9442	-.3721	13	1.4903	-.2602	38	1.2751	-.4586
14	.6541	.8625	39	1.6622	.1121	14	.8834	-.9224	39	2.2448	-.0400
15	.7031	.0227	40	.8079	-.4681	15	.9496	-.0416	40	1.0911	-.5421
16	.7503	.4434	41	.6534	-.4182	16	1.0133	.3746	41	.8824	-.4780
17	.8297	-.2930	42	.8031	-.1691	17	1.1206	-.3690	42	1.0846	-.2427
18	1.0674	.5441	43	1.3553	.1542	18	1.4416	.4463	43	1.8304	.0301
19	1.2988	.4886	44	.5131	.4672	19	1.7541	.3696	44	.6930	.4202
20	1.5194	-.0011	45	.8913	.3552	20	2.0520	-.1402	45	1.2037	.2735
21	1.2480	-.0026	46	.9786	.2645	21	1.6855	-.1370	46	1.3217	.1749
22	.6973	.2212	47	1.2354	.4091	22	.9417	.1573	47	1.6684	.2959
23	1.4352	-.2130	48	.9038	-.3921	23	1.9383	-.3444	48	1.2207	-.4749
24	.8283	-.0250	49	.7785	-.0207	24	1.1186	-.1008	49	1.0514	-.0920
25	.6760	-.6523	50	.9595	.1550	25	.9130	-.7142	50	1.2959	.0671
Moyenne				0.9821	0.0207					1.3263	-0.1037

Ce tableau montre aussi la différence entre les deux méthodes d'estimation des paramètres u et $1/\alpha$. Ces estimations étant asymptotiquement normalement distribuées, les estimations $1/\hat{\alpha}_1$ et $1/\hat{\alpha}_2$ rangées par valeurs croissantes ont été portées sur un graphique à loi de probabilité normale. Les deux distributions s'ajustent assez convenablement à une distribution normale, mais on voit que les estimations $1/\hat{\alpha}_2$ sont toujours au-dessus des estimations $1/\hat{\alpha}_1$. De plus, si les moyennes sont significativement différentes les pentes sont sensiblement égales.

Les estimations $\hat{u}_1$ et $\hat{u}_2$ s'ajustent aussi très bien à la distribution normale, mais cette fois sans différence significative ni en ce qui concerne les moyennes, ni les pentes (les graphiques n'ont pas été dessinés).

Ainsi, ces graphiques mettent en évidence le fait que la méthode modifiée donne lieu, pour les crues, à des lignes droites avec une pente plus grande que celle obtenue par la méthode classique des moments et conduit ainsi à une estimation plus prudente d'une crue future éventuelle. Pour confirmer ce résultat on a compté ensuite le nombre de fois où la plus grande valeur de chaque sous-échantillon était dépassée dans le reste de l'échantillon (Tableau 3, colonne 2). Prenant la plus grande valeur du premier sous-échantillon à partir duquel α et u étaient estimés par les deux méthodes, l'équation (7) a été utilisée et la plus grande valeur transformée correspondante a été calculée (Tableau 2).

Tableau 2

Les plus grandes valeurs et leurs transformations

Echantillon	Observations	Transformation Moments	Transformation Modifiées	Echantillon	Observations	Transformation Moments	Transformation Modifiées
1	2,135	1,741	2,280	26	2,749	2,598	3,403
2	1,927	1,284	1,757	27	2,826	2,771	3,598
3	3,688	3,916	5,125	28	2,116	1,129	1,624
4	2,934	2,462	3,225	29	3,113	3,525	4,565
5	2,249	1,706	2,264	30	3,275	3,309	4,539
6	2,779	2,512	3,349	31	1,388	0,977	1,234
7	3,445	3,530	4,665	32	3,620	4,413	5,938
8	2,516	2,485	3,366	33	3,656	4,673	6,119
9	2,845	3,460	4,561	34	3,735	4,229	5,560
10	2,746	2,369	3,119	35	3,602	4,254	5,654
11	2,362	2,558	3,182	36	4,054	4,826	6,447
12	3,806	4,965	6,602	37	2,513	2,266	2,918
13	2,795	2,925	3,905	38	2,190	1,695	2,333
14	0,991	-0,214	- 0,046	39	6,031	10,136	13,498
15	1,760	1,260	1,629	40	2,235	1,337	1,896
16	3,256	2,886	3,674	41	1,215	0,375	0,594
17	2,524	1,801	2,459	42	1,460	1,003	1,340
18	4,219	5,047	6,528	43	3,705	5,175	6,811
19	3,783	5,402	7,005	44	2,255	1,624	1,983
20	5,625	8,545	11,402	45	2,623	2,693	3,431
21	4,706	5,850	7,795	46	3,877	5,198	6,764
22	1,846	1,508	1,895	47	3,176	2,478	3,402
23	5,164	7,198	9,665	48	2,572	1,981	2,612
24	2,200	1,797	2,360	49	3,519	3,531	4,627
25	1,576	0,413	0,724	50	1,583	0,565	0,857

Tableau 3

Nombre des dépassements, observées et théoriques

Echantillon	Observation	Méthode des Moments	Modification
1	59	81	49
2	70	113	80
3	13	6	3
4	24	41	21
5	50	94	49
6	29	40	19
7	18	17	4
8	39	41	19
9	25	18	4
10	32	43	22
11	44	37	21
12	8	3	0
13	28	25	6
14	148	343	315
15	80	119	91
16	20	24	14
17	37	74	41
18	4	3	0
19	9	2	0
20	1	0	0
21	3	1	0
22	72	97	72
23	2	0	0
24	53	77	45
25	96	237	187
26	31	35	19
27	27	30	17
28	61	134	91
29	22	17	4
30	19	19	4
31	109	147	119
32	7	4	1
33	14	4	0
34	10	3	2
35	16	4	1
36	5	3	0
37	40	49	25
38	55	87	45
39	0	0	0
40	52	110	71
41	124	238	202
42	101	142	110
43	11	2	0
44	49	91	65
45	35	35	19
46	6	2	0
47	21	41	19
48	36	66	36
49	17	17	4
50	95	205	167
Somme observée	1,927	3,011	2,083
Somme théorique	2,225	---	---

Cette plus grande valeur a ensuite été comparée avec l'échantillon restant des 490 valeurs non transformées par l'équation (7), de manière à compter le nombre de fois où elle y était dépassée.

Cette procédure étant répétée pour chacun des 49 autres sous-échantillons : les résultats sont donnés dans les colonnes 3 et 4 du tableau 3.

Ainsi qu'il était escompté, la méthode modifiée, qui augmentait les valeurs les plus grandes d'une plus grande quantité, conduit à un plus petit nombre de dépassements que la méthode des moments, chaque dépassement étant interprété comme un ouvrage détruit.

L'espérance mathématique du nombre de dépassements pour la $m^{ième}$ plus grande valeur d'un sous-échantillon d'effectif n prélevé dans un échantillon de N + n est :

$$E(n,\ m,\ N) = \frac{mN}{n+1} \tag{11}$$

Dans notre exemple, $m = 1$, $N = 490$, $n = 10$. Par conséquent nous pouvons nous attendre à ce que la plus grande valeur d'un sous-échantillon d'effectif 10 soit dépassée en moyenne 44,5 fois dans un échantillon ultérieur des 490. Dans ces conditions le nombre théorique total de dépassements est égal à $50 \times 44,5 = 2225$.

La variance est donnée par :

$$\sigma^2(n,\ m,\ N) = \frac{Nm(n-m+1)(N+n+1)}{(n+1)^2(n+2)} \tag{12}$$

soit :

$$s(10,\ 1,\ 490) = 41,1 \tag{13}$$

Bien que les deux totaux sont à l'intérieur des limites acceptables de confiance, la méthode modifiée conduit à un total plus voisin de son espérance mathématique et donne un moindre nombre de dépassements dans chacun des 50 sous-échantillons. Ainsi la méthode modifiée d'estimation, conduisant à l'hypothèse d'une destruction de 2083 ouvrages au lieu de 3011 (soit 298 en moins) au cours de 500 années est "meilleure" (au sens défini précédemment) que la méthode classique des moments, puisqu'elle réduit les coûts de réparations et remplacements.

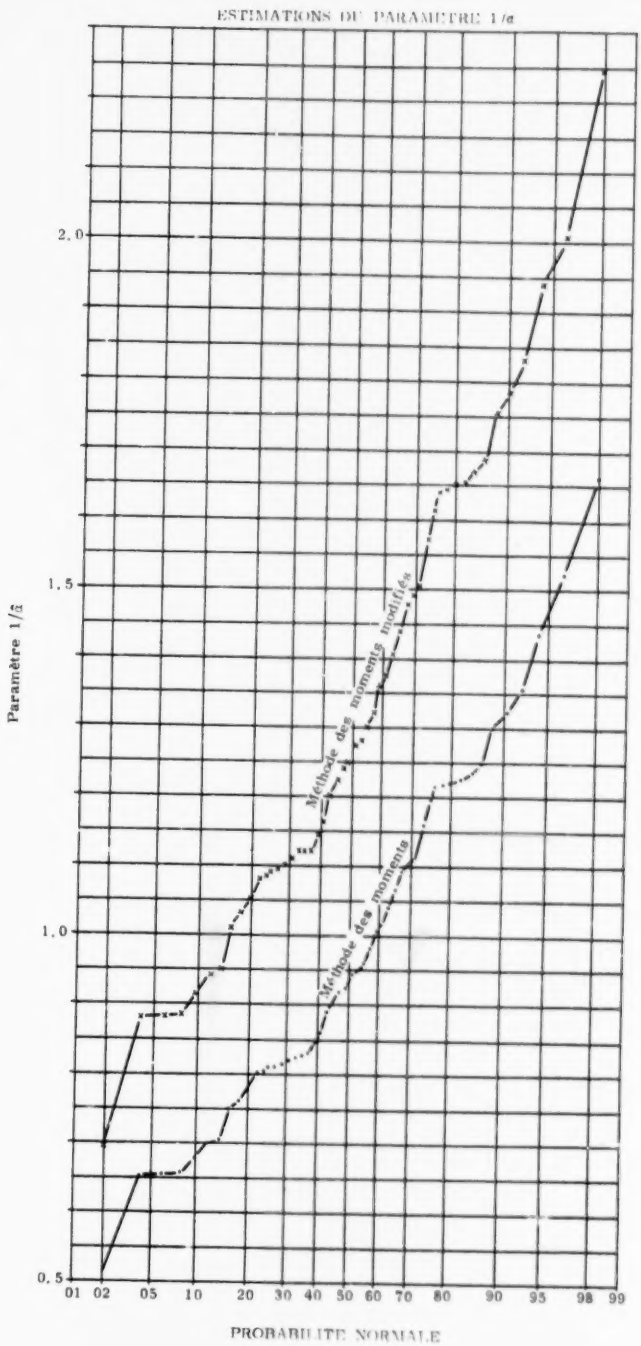
ESTIMATIONS DU PARAMETRE 1/a
Paramètre 1/a
2,0
1,5
1,0
0,5
Méthode des moments modifiés
Méthode des moments
01 02 05 10 20 30 40 50 60 70 80 90 95 98 99
PROBABILITE NORMALE

Figure 1

BIBLIOGRAPHIE

[1] BENSON M. (1962) - Plotting positions and economics of engineering planning, Proc. Amer. Soc. Civ. Eng. Paper 3317.

[2] LIEBLEIN J. (1954) - A New Method of Analyzing Extreme-value Data, N. A. C. A., Technical note 3053, Washington, D. C.

[3] GUMBEL E. J. (1954) - Statistical Theory of Extreme Values and Some Practical Applications, National Bureau of Standards Applied Mathematics Series No. 33, Washington, D. C.

[4] GUMBEL E. J. (1960) - Statistics of Extremes (Columbia University Press, New York).

[5] GOLDSTEIN N. (1963) - Random Numbers from the Extreme-value Distributions, Publ. de l'Institut. Statist. de l'Univ. de Paris, Vol. XII, 2, pp. 137 - 158.